DOCUMENT RESUME

ED 069 699 TM 002 150

AUTHOR	Fremer, John; And Others
TITLE	Student Involvement in Test Development.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-TDR-72-3
PUB DATE	May 72
NOTE	39p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; College Students; *Educational Improvement; Interviews; *Measurement Techniques; *Physical Education; Questionnaires; Rating Scales; Student Evaluation; *Student Participation; *Test Construction; Testing; Tests
IDENTIFIERS	Advanced Placement Studio Art Examination; College Board Achievement Tests in Mathematics; Cooperative Tests and Services Health Tests; National Assessment of Educational Progress; Undergraduate Program Physical Education Test

ABSTRACT

This report reviews the issue of student involvement in test development and presents summaries of instances of student contributions to tests and testing programs. The report goes on to describe a study in which a preliminary version of the Undergraduate Program Physical Education Test was administered on an experimental basis to a group of students majoring in physical education. These students evaluated a number of aspects of the draft test via a questionnaire and provided further reactions in interviews conducted by the authors. The responses of the students are analyzed and general themes identified. Suggestions are offered regarding future attempts to involve students in the test development process. (Author)

7/17/72

TEST DEVELOPMENT REPORT

TDR-72-3 MAY 1972

ED 069699

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

STUDENT INVOLVEMENT IN TEST DEVELOPMENT:

John Fremer
William Kastrinos
Chancey Jones

Test Development Division, ETS

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

ETS

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

TM 002 150

Copyright © 1972 by Educational Testing Service. All rights reserved.

ED 069699

STUDENT INVOLVEMENT IN TEST DEVELOPMENT:

JOHN FREMER
WILLIAM KASTRINOS
CHANCEY JONES
Test Development Division, ETS

Test Development Report
TDR-72-3
May 1972

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey

TABLE OF CONTENTS

Page

Abstract i

Introduction 1

Role of the Student in Testing 2

Recent History of Student Involvement in

ETS Test Development 5

Advanced Placement 5

College Board Achievement 6

Cooperative Tests and Services 7

National Assessment 7

National Teacher Examinations 7

Multiprogram Involvement of Summer

College Students 8

Climate for Student Involvement 8

Initiation of the UP Physical Education Test

Student Review Project 9

Administration of the Preliminary Test 10

Interview Procedures at Cortland 11

Summary of Student Review Sheets 13

Summary of General Comments by Students During Interviews 16

Specific Comments Made by Students During Interviews 18

Summary and Uses of Students' Ratings of Questions 19

Relationship Between Selection of Item for Comment and Item Difficulty . 21

Changes Related to Student Comments 22

If We Did It Again 22

Appendix A 25

Appendix B 26

Appendix C 35

Abstract

This report reviews the issue of student involvement in test development and presents summaries of instances of student contributions to tests and testing programs. The report goes on to describe a study in which a preliminary version of the Undergraduate Program Physical Education Test was administered on an experimental basis to a group of students majoring in physical education. These students evaluated a number of aspects of the draft test via a questionnaire and provided further reactions in interviews conducted by the authors. The responses of students are analyzed and general themes identified. Suggestions are offered regarding future attempts to involve students in the test development process.

STUDENT INVOLVEMENT IN TEST DEVELOPMENT: A CASE STUDY

INTRODUCTION

This report summarizes an effort to involve students in the development of a form of the Undergraduate Program[1] Physical Education Test. In addition, the report discusses the general issue of student participation in test development and summarizes a number of examples of student involvement in test development at ETS.

Physical education has not been thought of generally as a prime area for student protest, as compared, for example, with psychology and sociology, the chosen fields of many student activists. The effort to involve students in test construction was, indeed, not a response to demands by physical education students for representation in the development process, but rather an outgrowth of the concern on the part of the Program Directors and Test Development Division staff associated with the Undergraduate Program that student involvement be fostered.

The primary purpose of this project was to obtain from students majoring in physical education their reactions to (1) individual questions and (2) the test as a whole. There was no a priori commitment to changing the test or test specifications, but the comments of students were to be given careful consideration. Although it was realized that students from a single school could not be a determining

---

[1]The Undergraduate Program for Counseling and Evaluation offers tests for measuring the academic abilities and achievement of college students. The examinations are available on an institutional basis for enrolled students and are designed to provide reliable information for counseling and evaluation rather than for use in admissions.

force in changing specifications, it was felt that an attempt to involve students through an interview approach could be educational for the participating staff members and of value to the test being developed.

ROLE OF THE STUDENT IN TESTING

In many of the testing programs at Educational Testing Service, we use the term "test-taker" almost interchangeably with the word "student." This usage reflects a point of view that has a marked influence on our approach to the test development process. We devote considerable attention to the problems of reducing ambiguity and "trickiness" in our test questions so that students will be spared the frustration of attempting to second-guess our meaning and intentions. We try, in our test directions and in each of our questions, to communicate clearly to the student. Since there is a steady flow of information from ETS to the student, we try to monitor this flow very carefully.

There is a sign in Trenton, New Jersey, that proclaims "Trenton Makes, the World Takes." There is no sign at ETS suggesting that "ETS Makes (tests, that is), the Students Take," but no such sign is necessary. The basic model for ETS-student relationships assigns an active role to ETS and a passive one to students.

What about the flow of information to ETS from students? What do we learn from students? Or, perhaps more important, what should we be learning? We will explore the twin issues of what should be and what has been happening before going on to describe in detail the approach used in this project.

It takes only brief consideration of the relationship between testing and students to identify a fundamental inequity. Many of the significant testing experiences that a student undergoes were not designed primarily to serve the student, even though many students are helped to make educational decisions. Testing programs are most often planned and controlled by representatives of educational institutions that need to make critical decisions about students. The

admissions testing programs of the College Board, the testing done by the American College Testing Program, and tests by other professional organizations are perhaps the best-known examples of such test uses. In each of these selection programs, the student spends time taking tests so that an estimate of his developed ability can be derived for use by an institution. Since the institution typically will make a single "go or no-go" decision, the evaluation of the student's effort is reduced to but one number or a small set of numbers that can be treated in a prediction equation.

A respect for symmetry in relationships would seem to demand a greater respect for the needs of students in the design, development, and implementation of testing programs. This notion is brought into focus in the Report of the Commission on Tests wherein the Commission points out that the primary clientele of the College Board has not been students, but the admissions officers of member colleges, with guidance counselors and principals of secondary schools being the secondary clienteles. The Commission states "An emerging clientele of the College Board and one that should in the Commission's opinion be immediately adopted as a fully valued clientele, is composed of the students and adults out of school who are potentially entrants in programs offering post-secondary educational opportunities. Some of these potential entrants become involved in the Board's services now. As a result, they receive some information and supportive services from the College Board, but these are for the most part spun off from the services designed for admissions officers and are provided incidentally to meeting those officers' needs. Being served incidentally, the students are served less well and are essentially captive (and paying) customers rather than an equally valued clientele of the College Board."[1]

---

[1]Report of the Commission on Tests, Volume One, Righting the Balance, 1970, p. 56.

The programs that now exist may be efficient and highly appropriate mechanisms for serving institutions. How would they measure up, though, if we were to ask whether the student derives as much profit from his effort and his test fee as does the institution? One could argue that we cannot serve both the student and the institution. This position is especially easy to arrive at if we give lip service only to the goal of seeking equal treatment for students, all the while assuming that almost all of the current testing arrangements are unalterable.

One fruitful source of ideas regarding possible changes in current procedures is the involvement of a group that has no vested interest in maintaining the status quo in testing, indeed a group which tends to view the status quo in all areas with deep suspicion. This group has, moreover, an intense interest in the world of testing. We are referring, of course, to students, a group that could bring its first-hand experience and unique perspective to bear on the issue of appropriate ways of making testing more responsive to student needs. Some crucial issues, such as what a question actually is communicating to students, can be answered only by students. In addition to providing kinds of input that can come only from them, students can also supply guidance of the kind usually sought only from professional educators. Committees of Examiners study items for logical, grammatical, and content flaws. Committee members bring years of professional training and a high sense of commitment to this task. Some of the kinds of information provided by committees, though, could also be provided by students, and students seem quite willing, indeed even anxious, to help. This willingness might fade if many, many requests for advice came to the same students, but this turn of events seems unlikely. Students still seem to be quite strongly identified with receiving information and performing for evaluation, rather than with being seen as appropriate candidates for reviewing, or advising about, the work being carried out by others in behalf of students.

RECENT HISTORY OF STUDENT INVOLVEMENT IN ETS TEST DEVELOPMENT

Given the arguments in favor of student participation, to what extent have students been involved in test development at ETS? In an attempt to answer this question, a search was made for memoranda and reports in the files of the authors and other Test Development Division staff members including all Department Chairmen. In addition, a number of staff members in other parts of ETS were contacted concerning student involvement in ETS programs and projects.

Our initial goal was to compile, over a wide time span, an exhaustive list of examples of student involvement in ETS test development procedures. It soon became clear, however, that the effort would be greatly hampered by the incomplete nature of files that were never originally focused on this particular issue and by the fallibility of human memory. Our tentative historical efforts did alert us to the fact that ETSers have worked with students in the past but not on a regular or widespread basis. We found that recent history provided a substantial set of examples of student involvement, enough of a sample, in our judgment, to characterize the current situation at ETS. Some of these instances will be reported below. Even though we cannot support our impressions with counts of instances, it does seem clear that student involvement in test development at ETS has increased dramatically during the past two or three years.

Advanced Placement -- Students played a major role in the development of the Advanced Placement Studio Art Examination. Four students met for a full day with the Committee of Examiners for the examination. The students received draft copies of the course description and examination. They were asked to indicate what parts of the description they could not understand or accept, and they were asked to explain what kinds of art works they would have created to satisfy the requirements of the evaluation.

Another way that students have been involved in the Advanced Placement Program is as members of panels reporting at regular program conferences in the areas of Biology, English, and Mathematics. Such student panels were on the program at the Mathematics conferences in 1969, 1970, and 1971 and at the Biology conferences in 1968, 1969, 1970, and 1971. At the first three Biology conferences, students represented various high school and college grade levels. A 1971 panel for the Biology Test consisted of four students, one each from the freshman, sophomore, junior, and senior years of college. Students have reported on the value of the tests to them as individuals, as well as indicating how the test affected their course selection for college. They also commented on the content of the examinations, focusing on such factors as a comparison of the objective and essay portions. They mentioned the advantages of taking an Advanced Placement course instead of the first year course in college.

College Board Achievement -- A major example of student involvement originated with a request by the Committee of Examiners for College Board Achievement Tests in Mathematics. The committee asked ETS to go beyond the usual item analysis of new questions and to attempt to discover what a student actually thinks as he solves mathematical questions. The committee felt that the conjectures stimulated by item analyses should be checked occasionally against information obtained by the in-depth interviewing of candidates. At their April 1970 meeting, the Mathematics Committee proposed that a small-scale feasibility study be conducted to test the usefulness of this approach. Members of the TDD Mathematics Department designed the study. The committee designated a special Level II Mathematics pretest to be administered to appropriate groups of candidates. A total of 75 candidates from four high schools of varying characteristics in the Princeton, New Jersey, area took the pretest, and 15 of these students were selected for in-depth interviewing.

This study to assess the feasibility of supplementing the normative information of item analyses with the clinical information obtained from the in-depth interviewing of students is the subject of Test Development Memorandum 71-4.

Cooperative Tests and Services -- The development of the CTS Health Tests drew on the services of students for a contribution to the test development process that is frequently suggested in measurement textbooks, but to our knowledge rarely employed at ETS. The Committee of Examiners for the Health tests prepared 250 open-ended questions in the health area. A sample of one of these questions was the following: "What is the danger in taking marijuana?" The student responses were then used as a basis for preparing the options to the multiple-choice questions that comprised the final examination. To the extent possible, the language of the students was retained in the options for the final questions.

National Assessment -- Students at the high school, undergraduate, and graduate levels have contributed to ETS developmental work in a number of subject areas for the National Assessment of Educational Progress. During 1970, students contributed to the National Assessment of Writing, participating as members of a panel that included teachers and laymen. As members of the panel, the students were called upon to help interpret and elaborate specifications for National Assessment Writing exercises and to write prototype exercises. During 1971, the students, along with other contributors, developed exercises for the National Assessment of Writing exercise pool. In the fall of the year, students participated in conferences at which these exercises were reviewed.

National Teacher Examinations -- During the summer of 1970, three of the participants in the ETS Summer Program for Graduate Students in Measurement contributed to the development of the NTE Examination, Education in an Urban Setting. Members of the Committee of Examiners as well as various ETS staff members were

at that time reviewing items for the test. Three summer students; two blacks and one Mexican American who had expressed interest in the test, were given review copies of all materials and asked to make comments for consideration in the development process. One of the graduate students challenged the idea that a TDD staff member from a non-minority group had primary responsibility for the test. He felt this way even though the ETS staff members involved were working with a committee of minority-group members and had considerable support from minority-group ETS staff.

Two summer students carried out extensive reviews of the individual test questions and of the balance and coverage of the test. One student, a Mexican American, devoted considerable attention to the problem of identifying pejorative words and statements that might be offensive to minority-group members. In addition, this same student worked with ETS staff members to clarify a number of questions dealing with Mexican-American culture. The Committee of Examiners and ETS staff members made considerable use of the comments and suggestions made by the student reviewers as they developed the Education in an Urban Setting Examination.

Multiprogram Involvement of Summer College Students -- In addition to the involvement of students in specific projects for testing programs, students have held summer positions in three of TDD's departments over the past few years. As summer staff members, these students have contributed to the development of a considerable number of TDD tests. Several students have also participated in research on a number of aspects of the test development process.

Climate for Student Involvement -- Some indication of the sentiment which fostered the above-mentioned examples of student involvement and the instance reported in full herein can be seen in one of the recommendations in a report to the officers of ETS by a committee of staff members:

"Within Test Development pretesting, TDD staff members should be urged to carry on pretesting in ghetto schools or black colleges personally on every possible occasion, to discuss the pretest with minority group students, and to involve minority group faculties in reactions to the pretests."[1]

Although specifically addressed to minority/poverty student involvement, this statement reflects a growing feeling that students should be represented in the test development process.

INITIATION OF THE UP PHYSICAL EDUCATION TEST STUDENT REVIEW PROJECT

The possibility of involving students in the test development process for the Undergraduate Program Physical Education Test was first raised and approved at a joint Test Development-Program Direction planning session on October 21, 1970. It was indicated at this meeting that the chairman of the test committee would be willing to cooperate in a study that would involve the administration of a preliminary version of the test to students at her school, the State College of New York at Cortland. The full Committee of Examiners gave their support to the proposal and tentative plans were made to administer tests at Cortland College in December and to interview the students who had taken the tests soon thereafter.

Since no provision had been made in the test development schedule for a special "pretesting" administration of this nature, it was necessary to depart from normal TDD production procedures in order to obtain test copies. Arrangements were made, therefore, to prepare preliminary test copies from ordinary bond paper, rather than from planograph. The preliminary test was assembled from items approved at the meeting, and this test was edited, typed, proofread, revised, printed, and shipped to Cortland in time for an administration prior to December 10, 1970. This preliminary test also served as committee copy, i.e.; the members of the Committee of Examiners were asked to answer all the questions, to review

---

[1] Statement on Educational Testing and Minority/Poverty Needs: A Report to the ETS PEOPLE Committee, June 1970, Recommendation 12, p. 5.

each question for correcting possible ambiguities, etc., and to suggest possible revisions.

ADMINISTRATION OF THE PRELIMINARY TEST

Since the primary focus of the project was to obtain detailed reactions of students, it seemed desirable to limit the number of students tested to a group that could be interviewed by ETS staff members. A total of 43 students majoring in Physical Education and enrolled in senior year courses volunteered for testing and subsequent interviewing. The distribution of students tested by grade level and sex was as follows:

Junior Year	9	Males	13
Senior Year	32	Females	27
Graduate Students	2	Not Disclosed	3
	43		43

Teachers at Cortland administered the 150-item preliminary test. Plans called for the test to be administered with a two-hour time limit, the same time limit used in the Undergraduate Program. The teachers indicated to the students that one purpose of administering the test was to obtain an evaluation of the test by students. They also indicated, however, that the scores would be used by the school. This latter announcement had the general goal of maintaining a sense of seriousness about test performance. Students were asked to circle on the answer sheet the number of any item about which they would like to comment in a subsequent interview. In addition, after the students had finished the test, each student was asked to complete a Student Review Sheet containing eight questions, questions that are similar to those asked of faculty members who request inspection copies of UP tests. Appendix A is an example of a Student Review Sheet.

INTERVIEW PROCEDURES AT CORTLAND

On December 15, 1970, the authors of this paper interviewed 36 of the 43 students at Cortland who had taken the preliminary version of the Physical Education Test. The students were interviewed in groups of 1 to 4 students, each group working with one of the interviewers. A total of five interview hours were scheduled throughout the day, and the 4 to 10 students who came to each session were each assigned to one of the interviewers. One interview session was conducted in a large room in which each interviewer worked in a separate section; but for the remaining sessions, separate rooms were available for each group. At the start of each interview session, students were given their answer sheets, on which most students had circled some question numbers, and a copy of the preliminary form of the test.

Each group interview started out with a general discussion of the test, and then turned to two other major components. The first was a comprehensive discussion of all questions that the students wanted to comment upon, usually those questions circled by students on their answer sheets during the examination. The second was the rating of the questions in a particular section of the test on the following three point scale:

G = Good question, especially appropriate for use at Cortland

A = Appropriate and acceptable for use at Cortland

NA = Not acceptable or not appropriate for use at Cortland

In order to permit enough time for detailed consideration of the questions that stimulated comments by, and discussion among, the students, each interviewer focused attention on 50 of the 150 questions in the test.

The nature of the sessions varied, as might be expected, according to the temperament and interest of the group of students present. Although all students

participated in the discussion, some very actively and forcefully, a small number, perhaps 5 to 7, made only a few comments. Some of the group had time to comment on more than the 50 questions assigned to their group. The degree of participation of students did not seem to be related to their scores on the test; relatively low-scoring students contributed to the discussion as did their high-scoring colleagues.

SUMMARY OF STUDENT REVIEW SHEETS

A total of 36 students completed Student Review Sheets. The comments made by all students to each question were collated and are listed in Appendix B. In this section of the report, we present our intepretation of and reaction to the comments.

Question 1: Do you expect to do as well on this examination as other physical education majors in your institution?

The majority of the students answered "yes" to this question. However, this question seems pointless as now stated. It serves for most students as an occasion to indicate confidence or lack of it. Some students do comment on reasons why they might be at a disadvantage relative to other physical education students taking the test; e.g., "I have not taken tests and measurements." It might be more useful to pose a question like the following: "Do you feel that any aspect of your training to date; i.e., the courses you have or have not taken, would give you an advantage or disadvantage compared to other physical education majors at your institution?"

Question 2: Does this test fit the physical education curriculum of your college?

The majority of the students agreed that the test fits the curriculum of their college. The degree of agreement reflected in the comments seems extraordinary. There wasn't a single blanket "no." Also, the comments made by the students giving qualified yeses contained only two fairly specific comments, both relating to an emphasis at Cortland on motor learning and perceptual motor development. This unanimity of responses raises the question to what extent the results were related to the fact that the Chairwoman of the Women's Physical Education Department, Dr. Katherine Ley, developed the specifications

for the first Undergraduate Program Physical Education Test while serving as a consultant to the Undergraduate Program. The test that Cortland students took was not this first form, but a second form being developed one year later. The four-member Committee of Examiners for this second form, however, made only major changes in the specifications that had been developed by Dr. Ley.

Question 3: What areas of knowledge and abilities covered in the test are ones which you consider of the greatest importance in Physical Education?

This question uncovered considerable diversity among the physical education majors at Cortland. The number of Student Review Sheet comments regarding methodology and understanding individual needs was consistent with the oral reports of a large percentage of the students who were interviewed. (See the subsequent section, summarizing general comments made by students during interviews.) Some students listed more than one area. It might have been useful to have major areas listed and have the student check the area that he thought was of greatest importance.

Question 4: Are there some areas of knowledge or abilities which are not handled adequately in the test? What are they?

The responses would seem to indicate that the students were satisfied with the way the subject-matter was handled in the test with the possible exception of inadequate coverage of the area of application of knowledge. The answers to this question are not consistent with the answers to some of the other questions, and one has to speculate that time constraints did not permit the student to answer the questionnaire with the preciseness that we desired.

Question 5: Are there some things covered in the test which are overemphasized or relatively unimportant? What are they?

Most students felt some areas were overemphasized. The main criticisms of the students were directed toward the emphasis on the areas of tests and measurements, historical background (especially the questions related to knowing the names of leaders in the field), rules, and organizations. The responses to this item on the questionnaire agreed closely with the comments of the students during the interview.

Question 6: Generally speaking, does there appear to be an adequate balance between the testing of student's knowledge and the testing of his ability to apply the knowledge usefully? Please comment.

The majority of students felt that there was an adequate balance between knowledge and the ability to apply knowledge. About one-third of the students felt there were a disproportionate number of questions requiring specific knowledge.

Question 7: Is the level of performance expected of students in this test a reasonable one? On the average, are the questions either too elementary or too difficult to be of help in evaluating a student's progress?

The majority of students felt the level of the test was reasonable and this was consistent with their reactions during the interview. Very few of the students felt that the test was too difficult, and this is surprising considering that some of the students were juniors and had not taken courses in tests and measurements or courses on the organization and administration of physical education programs.

SUMMARY OF GENERAL COMMENTS BY STUDENTS DURING INTERVIEWS

The initial discussion in each interview group was focused on general comments about the preliminary form of the UP Physical Education Test. These comments were recorded for each interview group and then collated for all five interview groups by each interviewer. Finally, the comments recorded by all three interviewers were summarized to give a total picture of student reactions. As noted earlier, these comments support the statements made on the Student Review Sheets.

All comments made by more than one student are listed below, in an order corresponding to the frequency with which they were made; the most frequent comment is listed first.

1. Opinion Questions -- Most students condemned the use of questions that, in their judgment, depended on the opinion of the people writing a particular question. Quite a few students felt that they wanted to develop their own philosophy in certain areas or had developed philosophies that were contrary to the philosophies of their teachers and other experts. Some students felt that "value" or "opinion" questions would be fair if they were reworded to say, "In the opinion of most experts." These students reasoned that they should know the prevailing philosophy even if they disagreed with it. Other students felt that setting value questions in the context of the opinion of most educators would not help, because students will have to use their own teachers as a reference.

2. Length of Test -- Most students felt that there were too many questions covering too wide an area with too little time to answer them properly. They had specific objections to the fact that they were given the test in the late afternoon after they had attended classes. However, most of their objections were more general in nature. Some students felt that the most difficult questions should be located in the middle of the examination before the fatigue factor

became a problem. Many students indicated that fatigue had set in by about items 100-115, and several indicated that they felt that they were barely even reading the questions at the end of the test.

A number of lines of evidence suggest that the students were not given enough time to do the test. It is striking, therefore, that they scored as well as they did on an examination that is designed to be administered in two hours. An exact comparison of scores is not possible because the test was changed after the interviews, but in the opinion of the test development consultant, the changes made between the preliminary and final versions of the examination were such as to make the test slightly easier. Yet the 43 Cortland students obtained mean raw scores of 79 (S.D. of 18), whereas the students taking the final test during a later norming administration earned mean raw scores of 60 (S.D. of 22). Both sets of scores were corrected for guessing.

3. Tests and Measurements -- Most students felt that there were too many questions on tests and measurements. One student, for example, said that more than one-half of the test was on tests and measurements. (His estimate is quite exaggerated, but it does show what he felt about the test.) Some students did point out that they would not be taking tests and measurements until their senior year, but they still felt the topic was overstressed.

4. Trivial and Obscure Points -- Most students felt that it was inappropriate, on the one hand, to ask questions about facts that were so well known that everyone would know them even without taking courses in physical education. On the other hand, they felt it was equally inappropriate to ask questions about minor points in sports that are seldom played. The students indicated that such information could be obtained from source books whenever needed. In general, the students felt that material which could be obtained readily from source books should not be tested in a memory question.

5. Perceptual and Motor -- Several students felt that there should be more questions that involved the perceptual and motor analysis of activities.

6. Teaching -- Several students felt that the test should have a greater orientation toward methods of teaching.

7. Item Formats -- Several students expressed concern about the use of negative questions, questions using NOT, LEAST, and EXCEPT format, and questions in a format that allowed for a combination of statements being correct.

8. Women's Athletics -- Several students, both males and females, felt that questions on women's athletics posed problems for men. Some felt that the test was overbalanced in the area of women's athletics.

9. Tests and Teaching -- Several students felt that there was no relationship between scores on tests and ability to teach. They felt that tests measure only memory.

10. Ethics - Several students felt that there was not enough emphasis on ethical practices; e.g., not allowing an injured student to continue playing.

11. Kinesthesiology -- Several students expressed approval at the inclusion of kinesthesiology questions in the test.

SUMMARY OF SPECIFIC COMMENTS MADE BY STUDENTS DURING INTERVIEWS

After the initial general discussion, each group focused on 50 questions and made specific comments about each question. Most of the comments could be placed in one of four major categories as follows:

Based on opinion of individual (a value judgment)	41
Multiple answers	29
Trivial	14
Too easy	12

These frequencies indicate the number of times that at least one individual in an interview group made the indicated comment about a specific question. Often others in the group agreed with the comment, but there were occasions when other students in the group disagreed, sometimes vigorously, with the comment.

SUMMARY AND USES OF STUDENTS' RATINGS OF QUESTIONS

The number of students who rated each question varied because some groups were larger than others, and some groups rated only some of the 50 questions they were asked to pay particular attention to. Others rated their set of 50 and some additional questions. The ratings were used and analyzed in a number of ways. The use that contributed most directly to the development of the new form of the test was that of directing the test development consultant to questions that might be faulty. All questions that received several Not Acceptable or Not Appropriate ratings were carefully analyzed. The four questions that were subsequently dropped from the test as too trivial, for example, had the following pattern of ratings:

	Number of Ratings		
Question	G	A	NA
97	2	4	5
140	2	5	5
143	3	2	7
146	1	2	9

In order to interpret the ratings given to any particular set of questions, it will be useful to know that for the test as a whole the distribution of ratings was as follows:

G	A	NA
32%	41%	27%

As further background for interpreting the ratings, the average correlation among raters was computed, including all raters who had rated 25 or more questions. The quite small average correlation of .33 may well be somewhat of an overestimate of what would have been obtained if all the students had worked independently. Within each of the five groups, the students knew their colleague's ratings. The level of agreement among raters on specific items can be seen in Appendix C, which gives the distribution of ratings for each item.

One hypothesis about the ratings which received some attention was that students' ratings would be a function of their success on questions. It seemed possible that students would rate questions that they answered correctly as being Good (G) or Appropriate (A) and express their disapproval of questions they answered incorrectly by rating them Not Appropriate (NA). The distribution of ratings for questions answered correctly and for questions answered incorrectly was determined for the students who rated questions 1-50 and for the separate group that rated questions 51-100. The percentage of ratings in each category was as follows:

Rating Category	Per Cent of Correct Questions Placed in Category		Per Cent of Incorrect Questions Placed in Category	
	1-50	51-100	1-50	51-100
G	25%	29%	16%	17%
A	54%	52%	54%	35%
NA	21%	19%	30%	48%

Students in both groups had negative feelings about questions that they answered incorrectly, approximately 17% (i.e., 16% and 17%) of such questions received a rating of Good, whereas 27% (i.e., 25% and 29%) of the questions which were answered correctly were rated Good. The tendency to assign higher ratings to questions answered correctly is not as strong as might be predicted

given the fact that the students were holding their own answer sheets with the telltale red marks signifying wrong answers while they announced their ratings.

RELATIONSHIP BETWEEN SELECTION OF ITEM FOR COMMENT AND ITEM DIFFICULTY

In the preceding section, the relationship of question rating to question performance was studied. A similar analysis was performed for the questions for which students had circled the question number at the time that they took the test, indicating that they would like to comment on the question during the interview. Of the 43 students who took the test, a total of 36 circled some questions on their answer sheets. (The other students may have made mental notes but did not circle question numbers.) The mean number of questions circled was 14. The distribution of number of questions circled was as follows:

Interval	f
50	1
.	
.	
.	
26-30	3
21-25	7
16-20	6
11-15	10
6-10	3
1-5	6
0	7
	43

The relationship between circling and question performance can be seen in the following:

Per Cent Circled Items Answered Correctly -- 49%

Per Cent Uncircled Items Answered Correctly -- 63%

Students did show a greater tendency to want to comment on questions that they answered incorrectly.

CHANGES RELATED TO STUDENT'S COMMENTS

The preliminary form of the test was also reviewed by the members of the Committee of Examiners, and the committee comments were judged by the test development consultant along with the student comments before any revisions were made in the test. It is difficult, therefore, to give a precise accounting of the extent of changes attributable specifically to student comments. It appears, though, that the following changes were influenced by student comments:

6 questions -- Stems revised to add qualifiers that establish basis for selecting response

4 questions -- Dropped from test as "trivial"

4 questions -- Revised to make more precise or to reduce or eliminate ambiguity

3 questions -- Options changed to remove possibility of double key

3 questions -- Options changed to be more plausible

---

20 questions -- Total changes as a result of students' comments

Information provided by students as to areas of overemphasis, ambiguities, errors, and quality of questions obviously had some impact on the development of the test. It seems reasonable to expect, moreover, that the interview experience will have some small but continuing effect on the test development practice of the staff members who participated in the project and of other staff members who take note of what happened when we sought student assistance.

IF WE DID IT AGAIN

The interview project was a very productive one, and one that we all enjoyed very much. We were impressed with the willingness of Physical Education Department of the University of New York at Cortland who contributed to the project. We

welcomed the eagerness and enthusiasm of the students as they discussed the test with us and the high quality of their comments, criticisms, and suggestions. Despite our overall satisfaction with the way that the project developed, however, we feel that some of our procedures could easily be improved.

If we were to interview students who had taken a preliminary test form in the future, we would do the following:

1. Spend more time in planning the enterprise.
2. Decide in advance of our interviews in what way we will use the information obtained and collect it in a form most suited to that use.
3. Use a standard form to record judgments. At Cortland, one interviewer kept accurate records for each student, another fairly accurate records, and the third kept only group records. The third interviewer tabulated individual responses but did not associate the response with a particular student.
4. Use a simplified review form so that students would be faced with only one question at a time.
5. Explore the possibility of capturing some information with a tape recorder.
6. Allow more time in the schedule for the Committee of Examiners to react and make changes in the examination.
7. If possible, receive more detailed information from the host school concerning scheduling of interviews.

Despite the fact that the experience of interviewing students who had taken the test was, in our judgment, a valuable one, we recognize that there are many other possible ways of involving students. In our brief history of student involvement in ETS test development, we mentioned projects that called upon students

to report to policy boards and to Committees of Examiners and which asked students to review test questions and test specifications. We feel that each of these approaches can be effectively employed and that other useful techniques are available. What is essential is a commitment on the part of ETS as an organization and on the part of participating ETS staff as individuals to the principle that the group most affected by our tests have the opportunity to shape the way those tests are developed and used. We would expect a positive outcome from any procedure that is thoughtfully planned and scheduled. The planning should insure that students have a clearly defined task or set of tasks and that the students' contribution can be integrated with that of the Committee of Examiners and of ETS staff members.

APPENDIX A

SAMPLE STUDENT REVIEW SHEET

UNDERGRADUATE PROGRAM

PHYSICAL EDUCATION FIELD TEST

To the Reviewer: It will be of help to ETS in revising and improving the Undergraduate Program Field Test in Physical Education if you will respond to the following questions:

1. Do you expect to do as well on this examination as other physical education majors in your institution?_____

2. Does this test fit the physical education curriculum of your college? Please comment._____

3. What areas of knowledge and abilities covered in the test are ones which you consider of the greatest importance in Physical Education?

4. Are there some important areas of knowledge or abilities which are not handled adequately in the test? What are they?_____

5. Are there some things covered in the test which are overemphasized or relatively unimportant? What are they? _____

6. Generally speaking, does there appear to be an adequate balance between the testing of the student's knowledge and the testing of his ability to apply the knowledge usefully? Please comment._____

7. Is the level of performance expected of students in this test a reasonable one? On the average, are the questions either too elementary or too difficult to be of help in evaluating a student's progress?_____

8. Would you expect the students who do well on this test to be those who have demonstrated success in their course work in Physical Education?

General Comments (use other side for additional space)

Thank you.

APPENDIX B

SUMMARY OF STUDENT REVIEW SHEETS

Question 1: Do you expect to do as well on this examination as other physical education majors in your institution?

Responses:		
	Yes	25 students
	No	10 students
	Uncertain	1 student
	Total	36

Question 2: Does this test fit the physical education curriculum of your college? Please comment.

Responses:		
	Strong agreement	25 students
	Somewhat	9 students
	Don't know	1 student
	Special response*	1 student
	Total	36

*"It fits the curriculum, but not what is taught in the courses."

Question 3: What areas of knowledge and abilities covered in the test are ones which you consider of the greatest importance in Physical Education?

Responses:		
	Methodology	10 students
	Understanding individual needs	8 students
	Skills (practical application)	6 students
	Physiology of exercise	5 students
	Kinesiology, curriculum planning, test and measurements (3 students each)	(9 students total)
	Social and emotional, administration, setting up programs (2 students each)	(4 students total)

Philosophy, attitudes, scientific background, motor learning, teaching areas, activities for appropriate grades, theories, health, coaching, not sure what test is measuring, all areas important, blank

(1 student each) (12 students total)

Total 54**

**Some students indicated more than one area

Question 4: Are there some areas of knowledge or abilities which are not handled adequately in the test? What are they?

Responses:

Yes	17 students
No	11 students
Blank	8 students
Total	36

Areas listed by students who answered "Yes":

Application of knowledge 5 students

Student unrest-drugs-rebellion, new methods of teaching, teaching progression of skills, progressive movement in education, anatomy and physiology, sociology of sport, coaching, teaching areas, psychology, child's characteristics, specific situations, progressive education, how well a teacher can teach, child development, curriculum planning, motor development, specifics on men and women

(1 student each) 17 students

Total 22*

*Some students made more than one response.

Question 5: Are there some things covered in the test which are over-emphasized or relatively unimportant? What are they?

Responses:		
	Yes	28 students
	No	6 students
	Blank	2 students
	Total	36

Areas listed by students who answered "Yes":

Test and measurements	8 students
Historical aspects	5 students
Leaders in physical education, specific questions or trivia, opinion on running class-philosophy-value judgments	9 students
Curriculum, physiology of exercise, questions pertaining to women, scientific foundations, divisions of AAHPER and professional groups	10 students
Methods, terms related to skills	2 students
Total	34**

**Some students made more than one response.

Question 6: Generally speaking, does there appear to be an adequate balance between the testing of the student's knowledge and the testing of his ability to apply the knowledge usefully? Please comment.

Responses:		
	Yes	22 students
	No	6 students
	Comments without a Yes-No response	6 students
	Blank	2 students
	Total	36

Breakdown of 22 "Yes" responses:

No comment	10 students
There is a balance between knowledge and application	2 students
<u>But</u> many questions were ambiguous	2 students
However some of the application questions do not provide valid evaluation of one's understanding	2 students
Most questions were fair; must use your knowledge	1 student
You need both teaching methods and ability	1 student
I was challenged and had to really apply formal learning	1 student
But some questions depend on opinion	1 student
Questions made you think	1 student
But I haven't had some of the courses yet	<u>1</u> student
Total	22

Breakdown of 6 "No" responses:

Too much knowledge and factual, not enough practical and application	4 students
Too much curriculum, tests and measurements, and administration	1 student
Will explain in interview	<u>1</u> student
Total	6

Breakdown of 6 comments without a "Yes-No" response:

How can you test ability to apply knowledge without knowing the situation	1 student
Application of knowledge not tested very well	1 student
One's opinion affects the answering of some of these questions	1 student
Too many small trivia questions	1 student
There seems to be a greater interest in what a student can manage	1 student
Application easy if you know the material	1 student
Total	6

Question 7: Is the level of performance expected of students in this test a reasonable one? On the average, are the questions either too elementary or too difficult to be of help in evaluating a student's progress?

Responses:

Reasonable	21 students
Too elementary	3 students
They go from one extreme to another*	4 students
Too difficult	3 students
Hard to say	2 students
No answer	3 students
Total	36

*One student's comment -- "Some elementary and some require intelligent thought. Perhaps they could be combined and a medium found."

Question 8: Would you expect the students who do well on this test to be those who have demonstrated success in their course work in Physical Education?

Responses:		
	Yes	23 students
	No	5 students
	Not necessarily	4 students
	Depends on how you define success	2 students
	Possibly	2 students
	Total	36

Breakdown of 23 "Yes" responses:

No comment	17 students
But not necessarily those who are good teachers	4 students
Probably even the ones who just learned a little	
Possibly so. I'm not sure all the tests we take in Cortland PE are as relevant as this	2 students
Total	23

Breakdown of 5 "No" responses:

No comment	1 student
Some of the questions were common sense, and others ask for application of information I have not acquired	1 student
Those who do well will be those who regurgitate factual, preplanned, prelearned, prememorized knowledge	1 student
Test results are not always reliable for some who can't do well on standardized tests, yet do well in course work	1 student
Because the questions were based on opinion and memorization	1 student
Total	5

Breakdown of 4 "Not necessarily" responses:

No further comment	1 student
A lot of questions were common sense-- things I knew before taking the course	1 student
Many of the questions are dependent upon one's philosophy	1 student
Some people retain material for short time	1 student
Total	4

Breakdown of 2 challenges to definition:

Success in terms of what? Grades or becoming a better person in society by participating in physical activities	1 student
It depends on how you define success. Maybe better, maybe worse, it depends on the individual	1 student
Total	2

Question 9: Students' General Comments about the test:

Responses: Comments	13 students
No comment	23 students
Total	36

Analysis of responses:

A lot of questions were dependent on opinion	4 students
Too many NOT and LEAST questions	2 students
Student did not have some of the courses covering topics included in the examination	1 student
Better than a lot of PE tests, but there are a lot of personal philosophies being challenged	1 student
Test would be better in sections; one on activity or skill, another on foundations, education, and scientific principles	1 student
Some questions ambiguous; must know the specific situation	1 student
Not enough progressive educational thoughts employed in the test	1 student
Teaching areas could be handled better, not so much on testing	1 student
An A student might not be as good a teacher as a C student	1 student
Many of the questions are pure fact	1 student
Many questions too hard and had more than one key	1 student
Test is too long	1 student
The length of this test will, in my opinion, be a factor as to the validity or reliability of results. I took only 1 1/2 hours, and I feel I did not spend adequate time on some of the questions. At times I had to force myself to read through the whole questions. Realizing that there were students who completed the test in 50 minutes, I can't help but suspect that the length of the tests led them to skim through without any thought to questions or answers	1 student
Total	17*

*Some students made more than one comment.

APPENDIX C

SUMMARY OF STUDENTS' RATINGS OF INDIVIDUAL ITEMS[1]

G = Good A = Appropriate, Acceptable, Adequate NA = Not Adequate

	G	A	NA
1.	1	8	2
2.	1	10	0
3.	0	10	3
4.	5	6	0
5.	2	6	5
6.	1	7	3
7.	5	6	0
8.	4	7	0
9.	2	7	2
10.	5	8	0
11.	4	7	0
12.	4	7	2
13.	6	5	0
14.	2	7	2
15.	2	0	11
16.	1	9	3
17.	1	8	2
18.	1	6	4
19.	2	9	2
20.	3	4	4
21.	7	6	0
22.	1	4	6
23.	5	5	1
24.	3	8	2
25.	3	7	1
26.	1	2	7
27.	1	4	5
28.	1	5	2
29.	5	4	1
30.	1	6	6
31.	2	5	0
32.	0	5	3
33.	5	6	0
34.	0	5	2
35.	2	6	0
36.	2	1	6
37.	3	2	5
38.	1	2	7
39.	1	3	3
40.	2	4	0
41.	0	3	3
42.	0	1	4
43.	2	2	0
44.	2	1	0
45.	2	6	2
46.	2	5	0
47.	4	0	0
48.	2	2	0
49.	3	4	0
50.	0	2	3
51.	2	4	5
52.	5	6	0
53.	4	4	3
54.	3	5	3
55.	2	3	6
56.	6	5	0
57.	1	9	1
58.	3	4	3
59.	2	2	7
60.	6	5	0
61.	3	6	2
62.	7	2	2
63.	3	8	0
64.	4	6	1
65.	0	1	10
66.	1	6	4
67.	0	5	6
68.	4	3	4
69.	2	9	0
70.	1	9	1
71.	5	6	0
72.	2	9	0
73.	1	5	5
74.	1	1	9
75.	5	6	0
76.	2	9	0
77.	0	0	11
78.	2	8	1
79.	4	6	1
80.	1	8	2
81.	1	7	3
82.	3	4	4
83.	0	3	8
84.	2	9	0
85.	3	1	7
86.	0	3	8
87.	5	6	0
88.	5	5	1
89.	9	2	0
90.	4	5	2
91.	5	5	1
92.	0	0	11
93.	9	2	0
94.	0	0	11
95.	3	5	3
96.	9	2	0
97.	2	4	5
98.	0	1	10
99.	3	6	2
100.	0	2	9
101.	8	4	0
102.	7	4	1
103.	4	3	5
104.	4	3	5
105.	1	2	9
106.	7	5	0
107.	4	6	2
108.	4	3	5
109.	3	2	7
110.	2	5	5
111.	6	2	4
112.	3	9	0
113.	6	2	4
114.	4	3	5
115.	8	4	0
116.	3	6	3
117.	2	9	1
118.	8	3	1
119.	11	0	1
120.	7	3	2
121.	8	4	0
122.	5	2	3
123.	10	1	1
124.	11	1	0
125.	3	4	5
126.	2	7	3
127.	2	6	4
128.	4	4	4
129.	6	5	1
130.	4	4	4
131.	4	7	1
132.	5	6	1
133.	4	0	8
134.	7	1	4
135.	11	1	0
136.	8	4	0
137.	5	6	1
138.	7	4	1
139.	2	4	6
140.	2	5	5
141.	9	1	2
142.	4	3	5
143.	3	2	7
144.	6	5	1
145.	3	1	8
146.	1	2	9
147.	4	3	5
148.	5	3	4
149.	7	2	1
150.	9	3	0

[1] The variation in the number of responses reflects the fact that students singled out some questions for comments and ratings that were not included in the group of 50 questions that was analyzed systematically with their interview group.